Supplementary Material

1. Metrics of model-data comparisons

Root mean squared deviation

Root mean squared deviation (RMSD) between the simulated and proxy-based SST is used to described the model-data "consistence". In principle, low RMSD indicates good consistence, and vice versa. Standardized RMSD is simply defined as:

$$\text{RMSD} = \sqrt{\frac{\sum_1^n (SST\_dat_i - SST\_mod_i)^2}{n}} \qquad \text{(S1)}$$

SST_dat and SST_mod are proxy-based SST and model simulated SST respectively, and n is the number of proxy records.

Normalised RMSD

Due to the arbitrary atmospheric CO_2 concentration prescribed in the simulation, the simulation can be biased toward either a warmer or cooler climate. To this end, the normalised RMSD is introduced to get rid of this potential bias (Kennedy-Asser et al., 2019). The normalized RMSD is thus defined as:

$$\text{RMSD\_norm} = \sqrt{\frac{\sum_1^n ((SST\_dat_i - \overline{SST\_dat}) - (SST\_mod_i - \overline{SST\_mod}))^2}{n}} \qquad \text{(S2)}$$

$\overline{SST\_dat}$ and $\overline{SST\_mod}$ are proxy-based and model simulated SST averaged over the locations where proxy data are available.

RMSD matrix vary among different types of proxies and are overall large but of the same order of magnitude as the uncertainty of proxy-based SST estimates (defined as the 2σ deviations for $\delta^{18}O$, and as the range between 5% and 95% percentile SST estimates for TEX^{86}, Mg/Ca and Clumped isotope data, Table 2). RMSD matrix for the 55Ma-3x simulation are smaller than for the 55Ma-1.5x simulation (Table 2), suggesting the 55Ma-3x simulation is closer to proxy-based SST reconstruction. Only this simulation is evaluated in more details thereafter.

Comparing these two RMSD metrics, the normalised RMSD matrix shows slight improvement compared to the standardized RMSD (Table S1). Furthermore, the sign of the mean deviation between the model simulation SST and proxy-based SST indicates whether the model overestimates or underestimates SST. For instance, the positive sign of the mean deviation in $\delta^{18}O$ implies overestimates in model simulation for these $\delta^{18}O$ sites, whereas negative sign for TEX^{86}, Mg/Ca and clumped isotope imply underestimation. The sign of the potential biases (i.e. underestimates or overestimates) is in line with the different values of RMSD between annual and summer SST. For example, in an overestimated case (e.g. $\delta^{18}O$) the RMSD in annual mean is lower than in summer, whereas in an underestimated case (e.g. TEX^{86}) the RMSD is lower for summer (Table S1).

2. Benchmarks for model evaluation

To quantitatively estimate the simulation performance, two benchmarks are computed by following the methodology of Kennedy-Asser et al. (2019). Benchmark 1 assumes a uniform mean temperature, in which the mean temperature of all records for a given proxy is taken as a homogeneous value at all sites (Fig. S2b). Benchmark 2 is based on the assumption of the least squares linear fit through the proxy-based SST estimates with the cosine function of paleo-latitude from each-type-of-proxy sites, and is taken to yield a synthetic and latitudinally varying SST field (Fig. S2b). The best case is that model simulations perform better than both benchmarks, which is described as showing good performance as they are correctly modelling zonal and regional variation beyond this general latitudinal trend. The worst case is the simulations perform worse than both benchmarks, which means a poor performance and as they are failing to identify even the most basic variation in the dataset. If the simulations outperform the constant mean benchmark but not the latitudinal gradient benchmark, they are described as moderate good performance.

Table S1 shows the performance of the model simulation outperforms benchmark 1 across both RMSD metrics, but not benchmark 2 for some proxies. This means that the simulation outperforms the constant mean benchmark but not fully the latitudinal gradient benchmark, and thus it corresponds to the moderate good performance of Kennedy-Asser et al. (2019). Here we have used the different proxies separately, rather than combining them altogether as was done by Kennedy-Asser et al. (2019), due to their different values and uncertainty range.

References

Hollis, C. J., Dunkley Jones, T., Anagnostou, E., Bijl, P. K., Cramwinckel, M. J., Cui, Y., Dickens, G. R., Edgar, K. M., Eley, Y., Evans, D. and et al.: The DeepMIP contribution to PMIP4: methodologies for selection, compilation and analysis of latest Paleocene and early Eocene climate proxy data, incorporating version 0.1 of the DeepMIP database, Geosci Model Dev, 12(7), 3149–3206, doi:10.5194/gmd-12-3149-2019, 2019.

Kennedy-Asser, A. T., Lunt, D. J., Valdes, P. J., Ladant, J.-B., Frieling, J. and Lauretano, V.: Changes in the high latitude Southern Hemisphere through the Eocene-Oligocene Transition: a model-data comparison, Climate of the Past Discussions, doi:10.5194/cp-2019-112, 2019.

Table S1 Root-mean-square-deviation (RMSD) of simulated SST in the 55 Ma-3x simulation and proxy-based SST estimates (in °C), and comparisons with benchmarks for evaluation. The proxy-based SST estimates are from the DeepMIP dataset for the early Eocene (Hollis et al. 2019). Summer is defined as July-August-September in the Northern Hemisphere, and January-February-March in the Southern Hemisphere.

Type of proxy		**TEX^{86}**	**$\delta^{18}O$**	**Mg/Ca**	**Clum. isope**
Uncertainty range of proxy-data		15.1	3.4	6.7	5.1
RMSD	Summer	9.9	10.1	3.1	7.1
	Annual	13.7	7.5	6.4	5.3
Normalised RMSD	Summer	14.5	4.3	2.3	4.1
	Annual	8.9	6.8	4.4	5.0
Mean deviation (model-data)	Summer	-5.7	8.5	1.0	5.7
	Annual	-10.3	3.2	-4.4	-1.5
Benchmark 1	Summer	9.9	10.2	3.4	6.6
	Annual	14.2	8.1	6.8	7.6
Benchmark 2	Summer	8.7	9.0	2.1	6.0
	Annual	13.2	5.6	5.7	3.3

Table S2 Lookup table to convert BIOME4 megabiomes into ORCHIDEE Plant Functional Types (PFT). For each megabiome is attributed a combination of different PFTs, the sum of which must be equal to 100%. ORCHIDEE PFTs are as follows: Tropical Broadleaved Evergreen (TrBLE); Tropical Broadleaved Raingreen (TBLR); Temperate Needleleaved Evergreen (TNLE); Temperate Broadleaved Evergreen (TeBLE); Temperate Broadleaved Summergreen (TBLS); Boreal Needleleaved Evergreen (BNLE); Boreal Broadleaved Summergreen (BBLS); Boreal Needleleaved Summergreen (BNLS); Natural C3 grass (C3); Bare Ground (BG); Not Applicable (N.A.).

BIOME4 megabiome	ORCHIDEE PFT
Tropical Forest	90% TrBLE, 10% TBLR
Warm-temperate forest	30% TBLR, 30% TeBLE, 20% TNLE, 20% C3
Temperate forest	60% TBLS, 25% TNLE, 15% C3
Boreal forest	35% BNLE, 25% BNLS, 20% BBLS,10% TBLS,10% C3
Savannah and dry woodland (< 30°)	70% C3, 30% TBLR
Savannah and dry woodland (> 30°)	70% C3, 30% TeBLE
Grassland and dry shrubland (< 30°)	70% C3, 30% TBLR
Grassland and dry shrubland (> 30°)	65% C3, 20% BG, 15% TeBLE
Desert	90% BG, 5% TeBLE. 5% C3
Dry tundra	50% BG, 50% C3
Tundra	70% C3, 20% BBLS, 10% BG
Land ice	N.A.

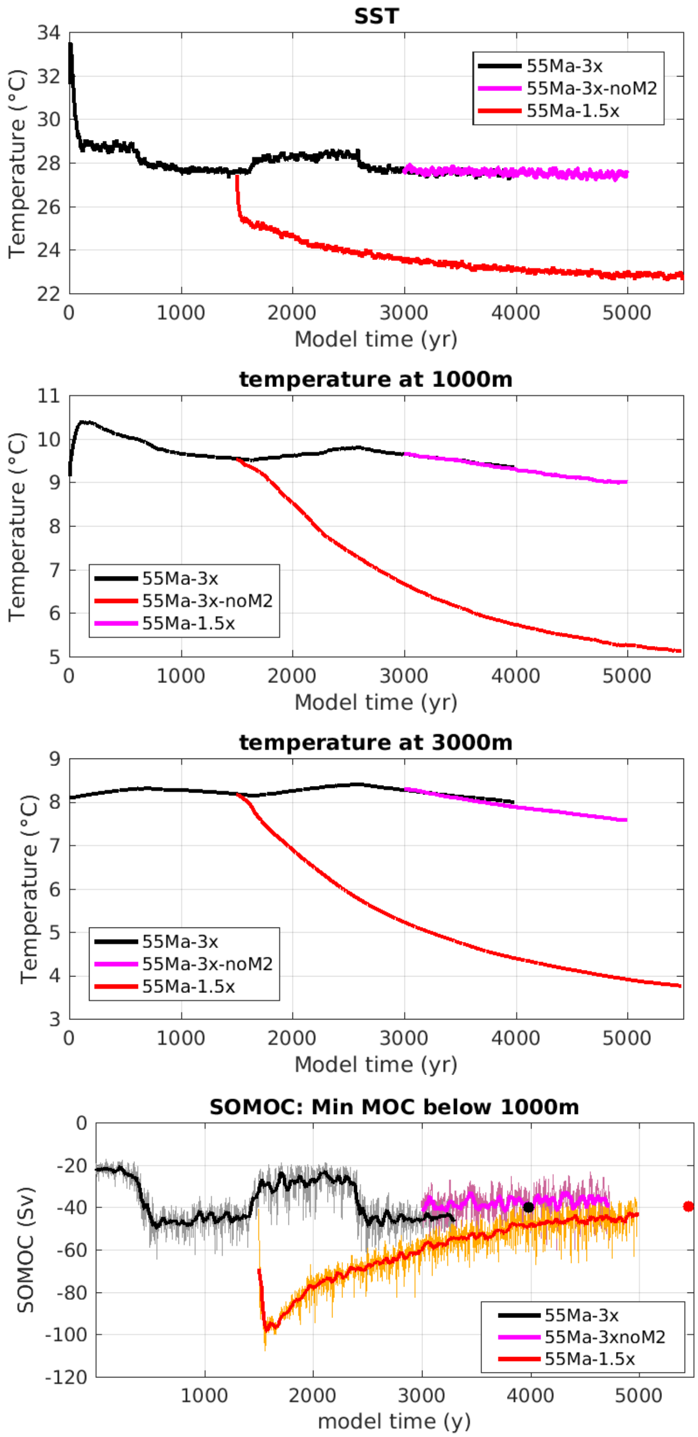

Figure S1. Time evolution of the global averaged temperature at the surface, 1000m and 3000m, and of Southern Ocean MOC index (defined as the minimum of the MOC below 1000 m) over the full integration of the 55 Ma-3x and 55 Ma-1.5x simulations. In the bottom panel, thin lines are for annual values, thick lines for 50-yr running means, and thick dots for the last 100-yr average.

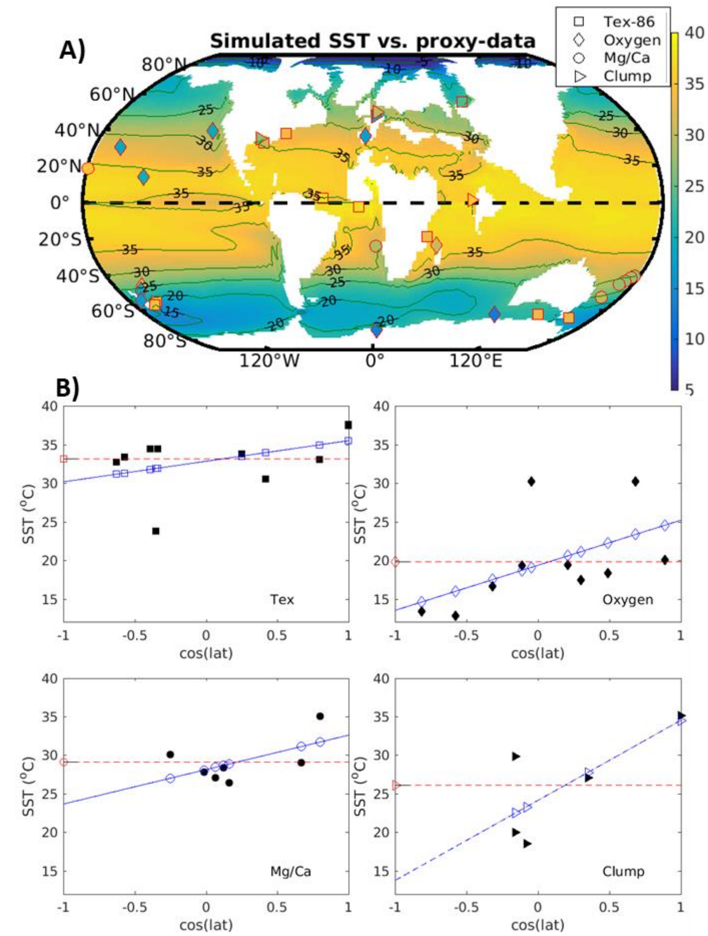

Figure S2 (A) Summer SST (in °C, averaged over July-August-September for the Northern Hemisphere and January-February-March for the Southern Hemisphere, the black dashed line indicates the discontinuity of the fields between the hemispheres) in the 55 Ma-3x simulation, and point-to-point comparisons with proxy-based SST estimates from the DeepMIP dataset for the early Eocene (Hollis et al. 2019). The different symbols represent different proxies. (B) Definition of the two benchmarks for each proxy: (1) the constant mean (in red), and (2) the latitudinal-gradient (in blue), computed from the available proxy-based SST (in black).

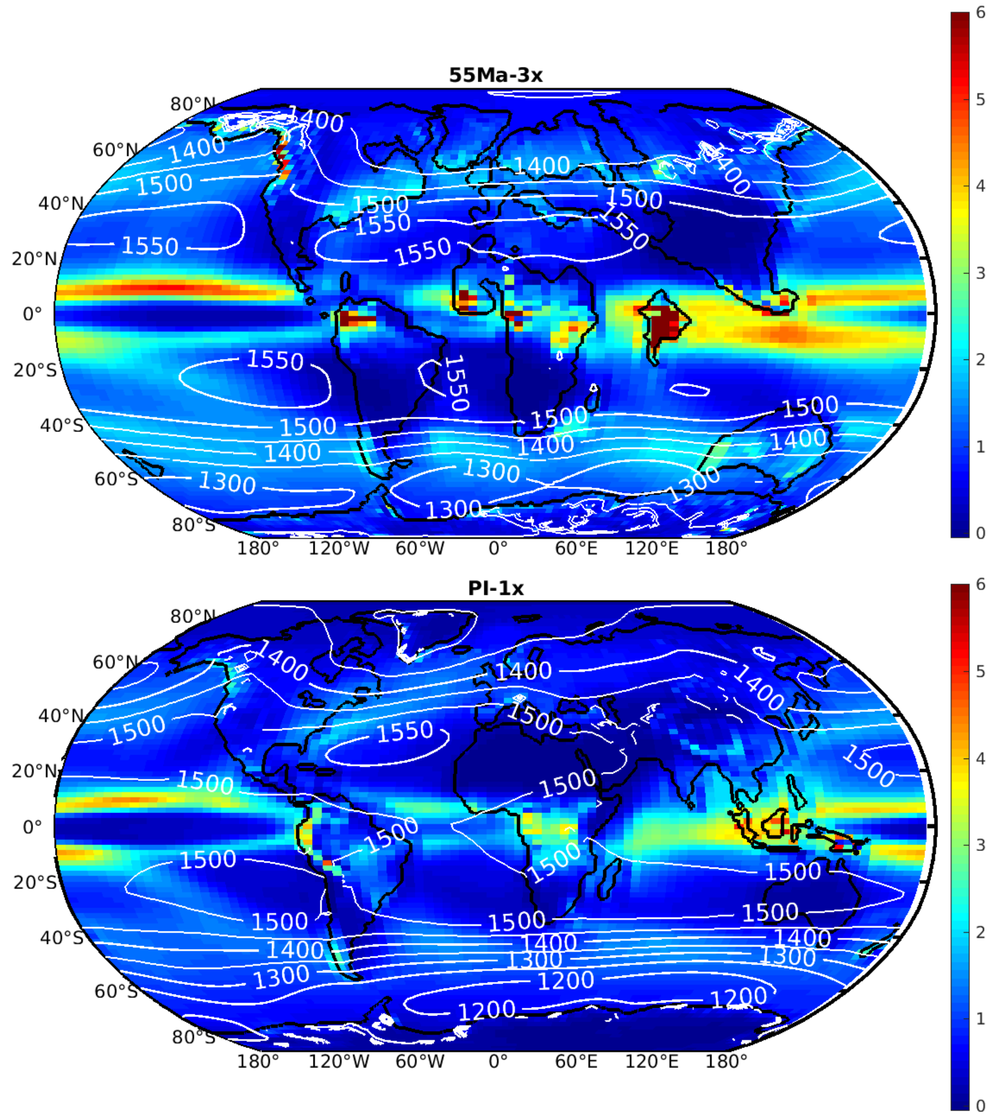

Figure S3 Geopotential height at 850mb (in unit of gpm, as contour) and precipitation (in unit of m/yr, as color) in the 55 Ma-3x (top) and in the PI-1x (bottom) simulations.

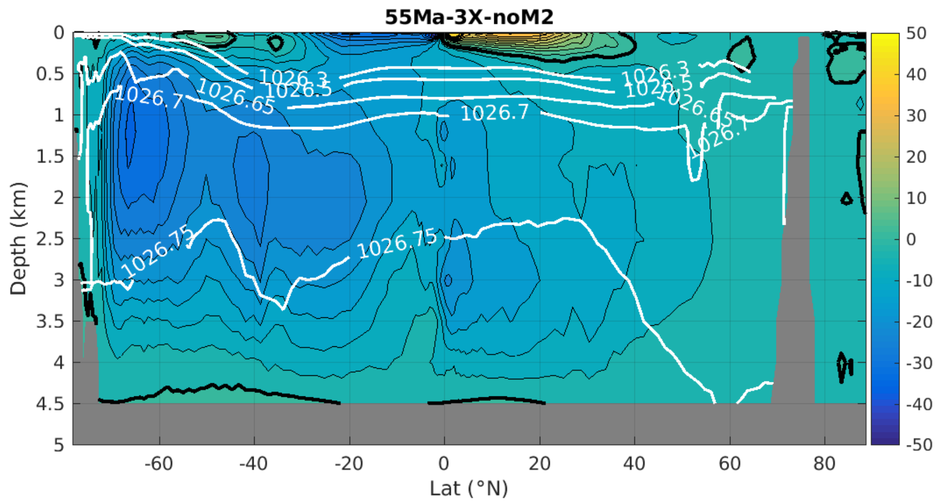

Figure S4 Same as Fig. 3 for the 55 Ma-3X simulation without tidal-induced mixing (55 Ma-3x-noM2).